

데이터분석과통계의이해 (2020-2) 중간고사

1. O/X 질문, X일 경우 간략한 설명을 적어주세요. (각 3점)

- A. (O/X) 파이썬은 컴파일 언어이다.
- B. (O/X) 파이썬의 변수는 레퍼런스이다.
- C. (O/X) 파이썬의 루프에서 생성한 변수는 지역변수이다.
- D. (O/X) Tuple은 수정이 불가능하다.
- E. (O/X) Dictionary의 key로 사용 가능한 데이터는 tuple, list, set이다.
- F. (O/X) Set은 다른 set을 원소로 가질 수 있다.

2. 숫자 n 을 받아, n 번째 피보나치 순열을 구하는 함수 $\text{fib}(n)$ 을 재귀적으로 구현하여라. (10점)

$$\begin{aligned}F_0 &= 1 \\F_1 &= 1 \\F_n &= F_{n-1} + F_{n-2}\end{aligned}$$

3. 피보나치 순열의 일반적인 재귀적 구현은 $\text{fib}(n)$ 계산에 약 n^2 번의 연산을 필요로 한다. 이는 $\text{fib}(n)$ 연산 시 이전 순열 값을 반복적으로 계산하기 때문이다. $\text{fib}(n)$ 연산 시, dictionary를 사용하여 순열의 값을 한번만 계산하도록 변경하여라. (5점)

Hint 1.: dictionary 생성 방법 $d = \text{dict}()$

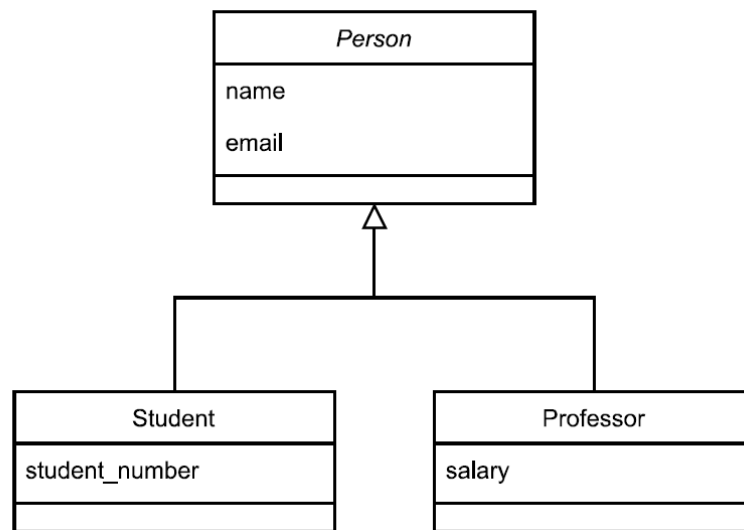
Hint 2. Dictionary의 주어진 key를 확인하는 법: $\text{key in } d \rightarrow \text{True or False}$

Hint 3: Dictionary의 주어진 key를 가진 값을 꺼내는 방법: $d[\text{key}]$

4. 다음 파이썬 프로그램의 출력 결과를 적으시오. (5점)

```
list1 = [1, 2, 3, [4,5,6]]
list2 = list1
print("#1")
print("l1", list1)
print("l2", list2)
list2[2] = 10
print("#2")
print("l1", list1)
print("l2", list2)
list1_copy = list1[:]
list1_copy[3][0] = 99
print("#3")
print("l1", list1)
print("l2", list2)
print("l3", list1_copy)
```

5. 다음의 클래스를 정의하여라. (5점)



사용 예:

```
person = Person('James', 'james@gmail.com')
student_1 = Student('John Mayer', 'jm@ajou.ac.kr', 20205888)
prof_1 = Professor('Kim', 'kim@ajou.ac.kr', 1000)
```

데이터분석과통계의이해 (2020-2) 중간고사

6. [10점] 머신러닝 알고리즘은 대표적으로 Supervised Learning과 Unsupervised Learning으로 구분된다. Supervised Learning과 Unsupervised Learning에 대해 간략히 기술하고, 대표적인 예시를 드시오. (Supervised Learning / Unsupervised Learning 설명 각 5점, 예시 각 3점)

7. 다음은 머신러닝 알고리즘과 관련한 주요 키워드이다. 각 키워드의 의미를 간략히 설명하시오.
(각 3점)

A. Hypothesis

B. Cost function

C. Gradient Descent

8. Stochastic Gradient Descent 알고리즘은 Gradient Descent 알고리즘이 가지고 있는 문제점을 개선하였다. 그 문제점에 대해 설명하시오. (5점)

9. Linear Regression에 관한 수식을 완성하시오.

n 개의 *input feature*, m 개의 입력 데이터를 가정

 $x_n^{(m)}$: m 번째 입력 데이터의 feature n $y^{(m)}$: m 번째 입력 데이터의 레이블 θ_n : 모델의 n 번째 파라미터

A. Hypothesis H (3점)

$H(\theta) =$

B. Cost Function J (3점)

$$J(\theta) = \frac{1}{2m} \boxed{}$$

C. Gradient Descent Algorithm (유도 과정 필요) (답만 적을 시 3점, 유도 할 경우 10점)

Repeat until convergence {

$$\theta_j := \theta_j - \alpha \frac{1}{m} \left[\begin{array}{c} \\ \\ \end{array} \right]$$

- $\alpha \stackrel{\text{L}}{=} \text{learning rate}$

10. Normal Equation

$X \in R^{m \times (n+1)}$: m 개의 입력데이터의 n 개의 feature

$Y \in R^m$: 크기 m 인 벡터: 입력데이터의 레이블

$\theta \in R^{(n+1)}$: 파라미터

A. Linear Regression의 Normal Equation $\theta = (X^T X)^{-1} X^T Y$ 을 유도하시오. (10점)

B. Normal Equation을 활용하여 다음의 데이터에 대해 Linear Regression 모델을 구하시오. (5점)

(x,y) : (1, 1), (2, 4), (3, 2)

